Theor Appl Genet (2003) 108:1–9
DOI 10.1007/s00122-003-1403-0

ORIGINAL PAPER

S. Yamanaka · E. Suzuki · M. Tanaka · Y. Takeda · J. A. Watanabe · K. N. Watanabe

Assessment of cytochrome P450 sequences offers a useful tool for determining genetic diversity in higher plant species

Received: 17 March 2003 / Accepted: 14 July 2003 / Published online: 13 September 2003

Abstract To investigate and develop new genetic tools for assessing genome-wide diversity in higher plant-species, polymorphisms of gene analogues of mammalian cytochrome P450 mono-oxygenases were studied. Data mining on *Arabidopsis thaliana* indicated that a small number of primer-sets derived from P450 genes could provide universal tools for the assessment of genome-wide genetic diversity in diverse plant species that do not have relevant genetic markers, or for which, there is no prior inheritance knowledge of inheritance traits. Results from PCR amplification of 51 plant species from 28 taxonomic families using P450 gene-primer sets suggested that there were at least several mammalian P450 gene mammalian-analogues in plants. Intra- and inter- specific variations were demonstrated following PCR amplifications of P450 analogue fragments, and this suggested that these would be effective genetic markers for the assessment of genetic diversity in plants. In addition, BLAST search analysis revealed that these amplified fragments possessed homologies to other genes and proteins in different plant varieties. We conclude that the sequence diversity of P450 gene-analogues in different plant species reflects the diversity of functional regions in the plant genome and is therefore an effective tool in functional genomic studies of plants.

Communicated by C. Möllers

S. Yamanaka · J. A. Watanabe · K. N. Watanabe (✉)
Gene Research Center, Institute of Biological Sciences,
University of Tsukuba,
1-1-1 Tennodai, 305-8572 Tsukuba, Japan
e-mail: nabechan@gene.tsukuba.ac.jp
Tel.: +81-29-8534663
Fax: +81-29-8534663

E. Suzuki · M. Tanaka · Y. Takeda · J. A. Watanabe · K. N. Watanabe
Faculty of Biology-Oriented Science and Technology,
Kinki University,
Uchita, 649-6493 Wakayama, Japan

E. Suzuki
Kihara Institute for Biological Research,
Yokohama City University,
Totsuka, 244-0813 Yokohama, Japan

M. Tanaka
Nara Institute of Science and Technology,
630-0101 Ikoma, Japan

Y. Takeda
Tokyo Rika Co. Ltd.,
Tokyo, Japan

Introduction

Both statistical analyses and the use of molecular markers are typically used in the examination of genetic diversity and/or species diversity at the molecular level. Analysis of plant genetic diversity has been conducted using many molecular markers such as isozymes, restriction fragment length polymorphisms (RFLPs), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphisms (AFLPs), simple sequence repeats (SSRs) and inter SSRs (ISSRs), etc. (Powell et al. 1996; Karp et al. 1997). Such markers usually record diversity in genetically neutral regions, but the question remains as to whether this diversity reflects the entire genome. This is a crucial consideration with respect to both genetic-database resource conservation (Karp 2002) and the monitoring of genetic variation in situ (Jarvis and Hodgkin 1999). We demonstrate in this study, however, that functional genomic analysis of multi-gene families, such as cytochrome P450, may overcome this problem.

Cytochrome P450 mono-oxygenases are widely found in animals, plants and micro-organisms (Shalk et al. 1999). In higher plants, cytochrome P450 (P450 or CYP) mono-oxygenases play important roles in oxidative detoxification and the biosynthesis of secondary metabolites (Kessmann et al. 1990; Donaldson and Luster 1991; Song et al. 1993; Teutsch et al. 1993; Ohkawa et al. 1998), and many P450 gene families have been found in various plant species.

Annotation of the *Arabidopsis thaliana* genome has shown that about 0.9% of 29,000 genes (272 genes and 26

Table 1 Plant species used in this study

Family	Species		Strain/variety
Solanaceae	*Lycopersicon esculentum*	Tomato	Ogata-tomato
			TA223
			Sugar lump
			Home-momotaro
			Legina
			TA209
	Capsicum annuum	Sweet pepper	Shishito
			Fushimi-kancho
			Kyo-midori
			Saitama-wase
	Solanum tuberosum	Potato	2x(v-2)7
			86.61.26
			F1-1
			S45-5
			T17-2
			Desiree
			Dejima
			Snowden
	Solamum acaule	Potato wild species	954.3CA
	Solanum brevidens	Potato wild species	CPC2451
	Solanum melongena	Eggplant	Ko-nasu
			seikokuchu-naganasu
			Canary-nasu
	Physalis alkekengi var. *franchetii*	Golden berry	Houzuki
Rosaceae	*Prunus armeniaca*	Apricot	
	Prunus avium	Cherry	
	Prunus persica	Peach	
	Prunus jamasakura	Cherry	Yama-zakura
	Rosa hybrida	Rose	
	Prunus yedoensis	Cherry	Somei-yoshino
	Kerria japonica	Japanese globe flower	
	Spiraea cantoniensis		
	Spiraea thunbergii		
Convolvulaceae	*Ipomoea batatas*	Sweet potato	Kokei-14
			Shiroyutaka
			Beniazuma
			Koganesengan
Libiatae	*Perilla frutescens* var. *crispa*	Perilla (blue)	
	Ocimum basilicum	Basil	
	Mentha piperita	Peppermint	
	Melissa officinalis	Lemonbarm	
	Perilla frutescens var. *crispa*	Perilla (red)	
Ericaceae	*Pieris japonica*	Japanese andromeda	
	Rhododendron scabrum		Hirado-tsutsuji
	Rhododendron obtusum		Kirishima
Legnminosae	*Lespedeza* spp.	Bush clover	
	Erythrina crista-galli		
Hippocastanaceae	*Aesculus turbinata*	Japanese horse chestnut	Japanese horse chestnut
	Ternstroemia gymnanthera		
Fagaceae	*Castanopsis* spp.	Chinquapin tree	
	Lithocarpus edulis		
Compositae	*Cichorium endivia*		
	Chrysanthemum frutescens		
Apocynaceae	*Catharanthus roseus*		
Ulmaceae	*Zelkova serrata*	Zelkova tree	
Platanaceae	*Platanus* spp.	Plane	
Thymelaeaceae	*Daphne odora*		
Moraceae	*Ficus carica*		
Begoniaceae	*Begonia* spp.		
Calycanthaceae	*Chimonanthus praecox*		
Lauraceae	*Cinnamomum camphora*	Camphor tree	
Araliaceae	*Hedera* spp.		

Table 1 (continued)

Family	Species	Strain/variety
Geraniaceae	*Geranium thunbergii*	
Magnoliaceae	*Magnolia grandiflora*	Magnolia
Berberidaceae	*Nandina domestica*	
Elaeagnaceae	*Elaeagnus* spp.	
Hydrangeaceae	*Hydrangea macrophylla*	
Malvaceae	*Hibiscus mutabilis*	Hibiscus
Cornaceae	*Aucuba japonica*	
Ginkgoaeae	*Ginkgo biloba*	Ginkgo
Dryopteridaceae	*Dryopteris* spp.	Fern
Cycadaceae	*Cycas revoluta*	Cycad

Table 2 Sequence information of PCR primers constructed based on the known P450 genes

Primer	Sequence (5′ to 3′)	Reference
CYP1A1F	*GCC AAG CTT TCT AAC AAT GC*	Inui et al. (2000)
CYP2B6F	*GAC TCT TGC TAC TCC TGG TT*	Inui et al. (2000)
CYP2C19F	*TCC TTG TGC TCT GTC TCT CA*	Inui et al. (2000)
CYP1A1R	*AAG GAC ATG CTC TGA CCA TT*	Inui et al. (2000)
CYP2B6R	*CGA ATA CAG AGC TGA TGA GT*	Inui et al. (2000)
CYP2C19R	*CCA TCG ATT CTT GGT GTT CT*	Inui et al. (2000)
Heme2B6	*ACC AAG ACA AAT CCG CTT CCC*	Kiyokawa et al. (1997)
Heme2C19	*TCC CAC ACA AAT CCG TTT TCC*	Kiyokawa et al. (1997)

pseudogenes) could be categorised as cytochrome P450 genes (http://arabidopsis-p450.biotec.uiuc.edu/; http://arabidopsis.org/info/genefamily/p450.html; http://drnelson.utmem.edu/CytochromeP450.html; http://www.biobase.dk/p450/; Riechmann et al. 2001). However, the functions of many of these annotated genes have not yet been demonstrated. Previous studies on plant cytochrome P450 genes mainly focused on individual gene functions and gene products, either within a species or between a small number of strains, but information concerning genetic diversity in different plant species still remains sparse. Furthermore, P450 genes that have so far been characterised are very diverse, being very variable in their gene alignments (Ohkawa et al. 1998). In contrast, however, our preliminary analysis of various monocotyledous and dicotyledous plant species by Southern hybridisation with rat and human *CYP1A1*s, suggested that analogues of mammalian *CYP1A1*s are common in angiosperms (Watanabe et al. 2000; Suzuki et al. 2001).

To further elucidate the above questions and to expand upon the use of genomics for analysis of genetic diversity in plants, we used the functional regions of the mammalian cytochrome P450 gene family to assess genome-wide diversity in a range of plant species. We describe our findings in this case report, in which our data have lead to the development of a novel application of functional genomics for the analysis of genetic diversity of model plant species.

Materials and methods

Plant materials and DNA extraction

Material from 51 plant species, representing 28 families, used in this study are listed in Table 1. Total DNA was extracted from each sample following the procedure of Wagner et al. (1987) and Bousquet et al. (1990). Polyphenol inhibition was caused by addition of polyvinyl pirroridone (K-30), 20 g/l to the extraction buffer.

Data mining in the *Arabidopsis* genome associated with mammalian P450 fragments

To examine the possibility of using mammalian cytochrome P450 genes as new genetic markers and indicators of plant genetic diversity, eight PCR primers including three forward primers (CYP1A1F, CYP2B6F and CYP2C19F) and five reverse primers (CYP1A1R, CYP2B6R, CYP2C19R, heme2B6 and heme2C19), were designed using human P450 sequences (Inui et al. 2000, Table 2). Two of five reverse primers, heme2B6 and heme2C19, were designed using sequence information from the heme-binding site of the human genes, *CYP2B6* and *CYP2C19*, respectively (Kiyokawa et al. 1997).

In preliminary studies, the distribution of these primer sequences in the *A. thaliana* genome was analysed, with minimum homologies set at 90%, using a BLAST-search based on sequence information from the *Arabidopsis* genome database [The Arabidopsis Genome Initiative (TAGI), http://www.arabidopsis.org./agi.html; AtDB's Arabidopsis Physical Map Overview, http://www.arabidopsis.org/cgi-bin/maps/Pchrom].

PCR amplification

In this study, primer-sets based on combinations of three forward primers (CYP1A1F, CYP2B6F and CYP2C19F) and five reverse

primers (CYP1A1R, CYP2B6R, CYP2C19R, heme2B6 and heme2C19) were utilised (Table 2).

PCR amplification was performed using 20 ng of extracted DNA in a total volume of 25 μl containing 1×PCR buffer (Qiagen), 0.16 mM of dNTPs, 1 μM of each primer and 1 unit of *Taq* polymerase (Qiagen). A total reaction of 32 cycles was programmed for 1 min at 94°C, 2 min at each annealing temperature and 3 min at 72°C in a Thermal Cycler (Applied Biosystems). Annealing temperatures for each primer-set are listed in Table 3. PCR products were electrophoresed in a 1% agarose gel.

Sequencing analysis and homology search

PCR products were cloned into a pGEM-T vector (Promega), and were sequenced on both strands, using the same primer-sets as used for PCR, using a BigDye Terminator Cycle Sequencing Kit with an ABI 377 sequencer (Applied Biosystems). Sequence data without primer sequence homology were further analysed using a BLAST search.

Results

Distribution of mammalian P450 primer sequences in the *Arabidopsis* genome

A total of 301 candidate DNA clones, which contain the expected annealing sites based on primer sequences (Table 2), were data-mined. From this sample set, 214 fragments were mapped to chromosomal locations within the *Arabidopsis* genome. The distribution of the eight primer sequences is shown in Fig. 1. These data indicate

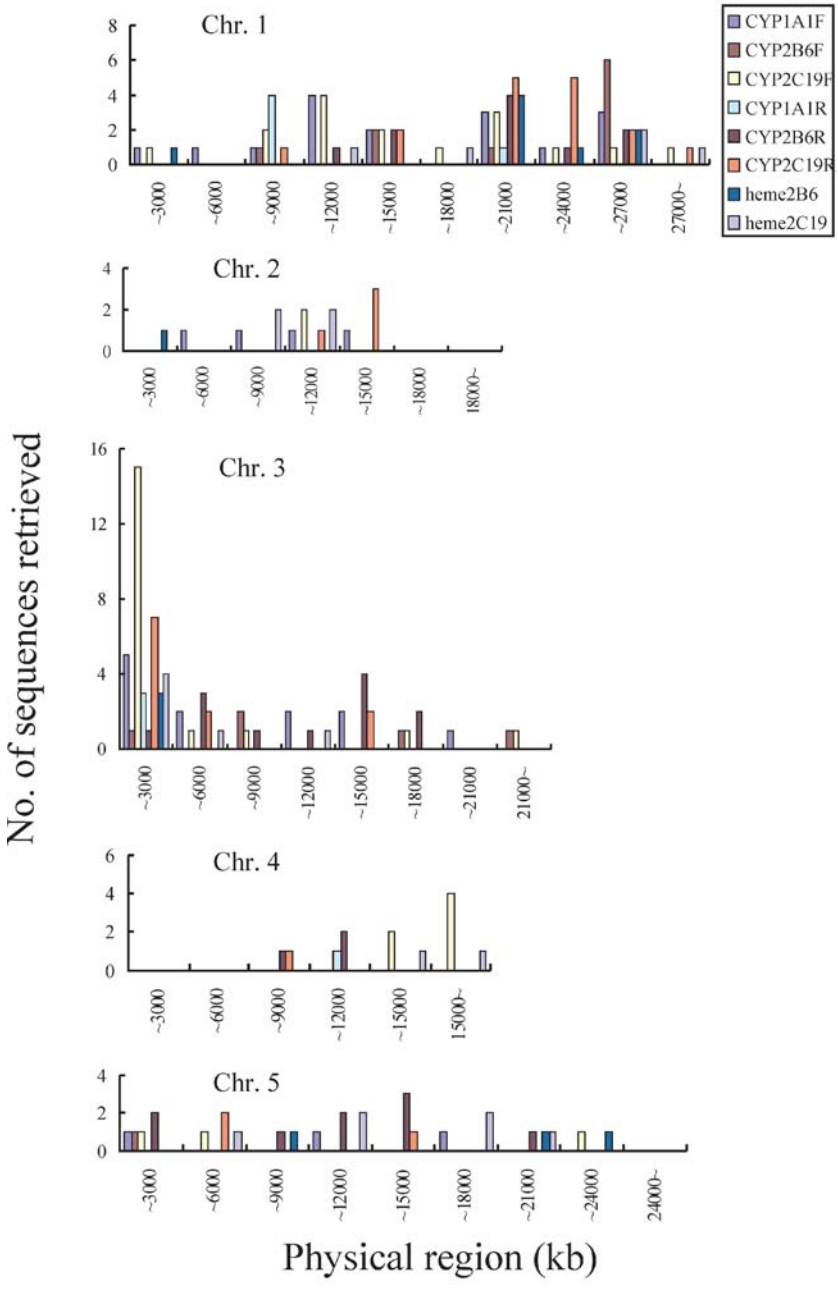

Fig. 1 Distribution within the *Arabidopsis* genome of sequences homologous to the eight primers derived from mammalian P450 genes. Each *coloured bar* indicates the number of retrieved sequences homologous to the indicated primer in the corresponding regions categorized within each 3,000-kb segment of the physical location

Table 3 Annealing temperatures of each primer-set and the total number of fragments scored from 51 species representing 28 families

Primer-set (forward/reverse)	Ann. temp.	No. of fragments
CYP1A1F/CYP1A1R	56.0	63
CYP1A1F/CYP2B6R	52.0	63
CYP1A1F/CYP2C19R	46.5	63
CYP1A1F/heme2B6	56.0	58
CYP1A1F/heme2C19	56.0	42
CYP2B6F/CYP1A1R	52.0	57
CYP2B6F/CYP2B6R	52.0	43
CYP2B6F/CYP2C19R	46.5	41
CYP2B6F/heme2B6	52.0	42
CYP2B6F/heme2C19	52.0	60
CYP2C19F/CYP1A1R	56.0	59
CYP2C19F/CYP2B6R	52.0	68
CYP2C19F/CYP2C19R	46.5	61
CYP2C19F/heme2B6	56.0	58
CYP2C19F/heme2C19	56.0	57

that primer sequences derived from the mammalian P450 genes are distributed throughout the *Arabidopsis* genome, suggesting the possibility of using them for an assessment of genetic diversity in higher plant species.

PCR amplification in diverse plant species

Every primer combination that was employed, successfully amplified fragments in each species used in this study and this suggested that P450 primers could be used in a wide range of plant species (Table 3, Fig. 2a). The total number of amplified fragments using each possible primer pair with each plant-species type ranged from 41 to 68 (Table 3). In addition to species-specific fragments, we obtained several family specific fragments by PCR. We observed the presence of fragments that were common to taxonomic families such as *Solanaceae*, *Rosaceae* and *Libiatae*. This suggested that there are both family specific and species-specific conserved regions in P450 genes and their analogues. It also indicated that these conserved regions could be effective genetic markers for many plant species, which would be particularly useful in cases where good sequence information is not available.

Plant genetic diversity estimated from P450 gene polymorphisms in *Solanaceae*

In *Solanaceae* species, every primer combination amplified distinct fragments (Table 4), suggesting that P450 primers could be used to identify species within this family (Fig. 2b). It was also observed that there were several species-specific fragments suggesting that polymorphisms exist that reflect genetic diversity within *Solanaceae*. This further indicates that P450 genes and their analogues are useful as genetic markers in plants. The high frequency of intra-specific polymorphisms, shown in Table 4, also showed that PBA (P450 based

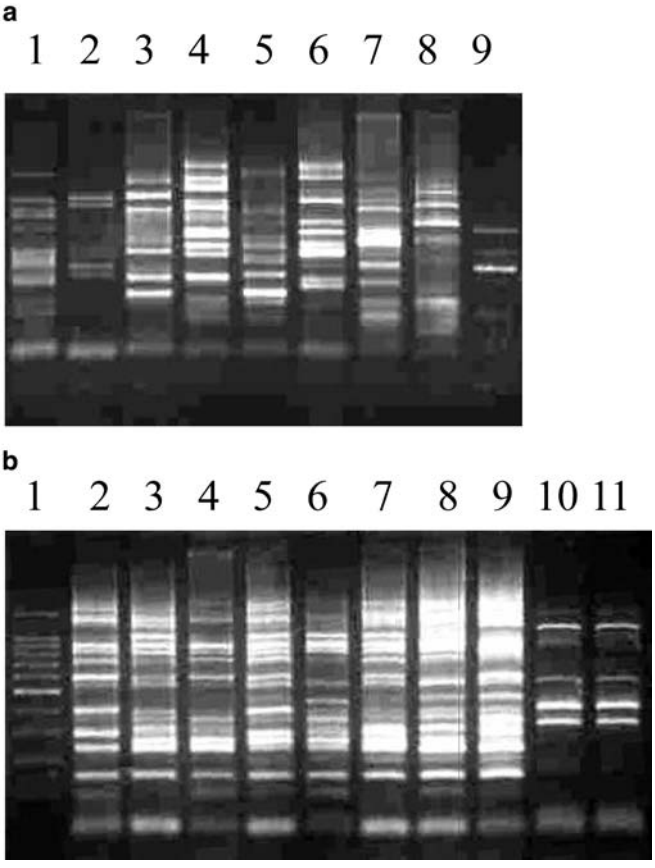

Fig. 2 a The results of PCR amplification with the primer-set, CYP2C19F/heme2C19 in diverse plant species. *Lane 1*:100-bp ladder, *2*: *Kerria japonica*, *3*: *Pieris japonica*, *4*: *Perilla frutescens* var. *crispa* (red perilla), *5*: *Castanopsis* spp., *6*: *Physalis alkekengi* var. *franchetii*, *7*: *Magnolia grandiflora*, *8*: *Hydrangea macrophylla*, *9*: *Cinnamomum camphora*. **b** The results of PCR amplification with the primer-set, CYP2C19F/CYP2C19R, in *Solanaceae* species. *Lane 1*:100-bp ladder, *2–9*: *S. tuberosum* [2×(v-2)7, Desiree, S45-5, Snowden, T17-2, Dejima, 86.61.26, F1-1], *10 and 11*: *L. esculentum* (TA223, TA209)

analogue) markers do reflect genetic diversity effectively. In Fig. 2b the length distribution of the amplified fragments was mainly below 1 kb. This was expected as the predicted fragment lengths from amplifications with *CYP1A1*, *CYP2B6* and *CYP2C19* were about 400, 600 and 750 bp, respectively. Our results were consistent with these expected fragment sizes and demonstrated that these primers could anneal to their target sites within the P450 genes, and their analogues and the reproducibly amplify fragments that were predicted from the *Arabidopsis* genome (Fig. 1).

Sequencing analysis of P450 gene-analogues

Fragments amplified with each primer-set showed homologies to a range of deduced protein sequences. Most fragments showed high homology to a variety of proteins found in different plant species (Figs. 3 and 4), suggesting

Table 4 Number of PCR fragments observed in *Solanaceae* and frequencies of intra-specific polymorphisms using mammalian P450 associated primers

Primer-set	Total fragments[a] (n=21)	Frequency of polymorphic fragments (%)			
		C. annuum (n=4)	*L. esculentum* (n=6)	*S. tuberosum* (n=8)	*S. melongena* (n=3)
CYP1A1F/CYP1A1R	38	22/29 (76)	22/23 (96)	24/25 (96)	18/21 (86)
CYP1A1F/CYP2B6R	33	16/18 (89)	26/26 (100)	19/19 (100)	8/8 (100)
CYP1A1F/CYP2C19R	35	12/17 (71)	15/17 (88)	22/22 (100)	17/19 (89)
CYP1A1F/heme2B6	45	21/23 (91)	26/26 (100)	38/38 (100)	18/24 (75)
CYP1A1F/heme2C19	32	15/18 (83)	20/20 (100)	14/19 (74)	21/25 (84)
CYP2B6F/CYP1A1R	32	17/19 (89)	22/24 (92)	21/24 (88)	13/18 (72)
CYP2B6F/CYP2B6R	36	23/26 (88)	23/23 (100)	26/26 (100)	10/15 (67)
CYP2B6F/CYP2C19R	26	13/18 (72)	22/22 (100)	15/15 (100)	8/8 (100)
CYP2B6F/heme2B6	37	23/24 (96)	25/25 (100)	15/20 (75)	6/14 (43)
CYP2B6F/heme2C19	36	15/15 (100)	23/23 (100)	22/23 (96)	13/13 (100)
CYP2C19F/CYP1A1R	44	31/32 (97)	32/32 (100)	35/35 (100)	5/11 (45)
CYP2C19F/CYP2B6R	56	26/29 (90)	34/35 (97)	26/27 (96)	17/20 (85)
CYP2C19F/CYP2C19R	48	15/18 (83)	16/18 (89)	27/29 (93)	13/21 (62)
CYP2C19F/heme2B6	37	19/23 (83)	31/31 (100)	21/25 (84)	7/14 (50)
CYP2C19F/heme2C19	36	21/21 (100)	20/20 (100)	29/29 (100)	4/10 (40)

[a] Total number of fragments observed in four species (21 strains or varieties)

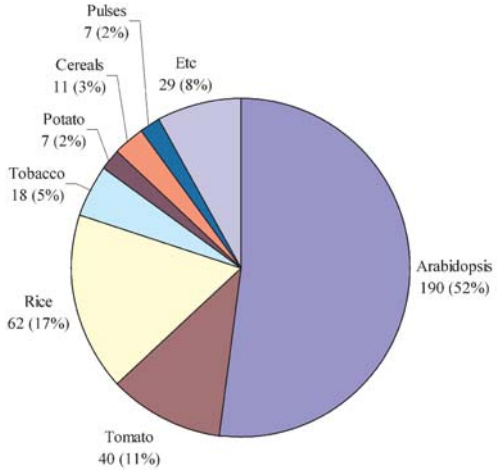

Fig. 3 Summary of a homology search of the indicated annotated regions using 365 fragments amplified from tomatoes with 15 PBA primer sets derived from mammalian P450 genes

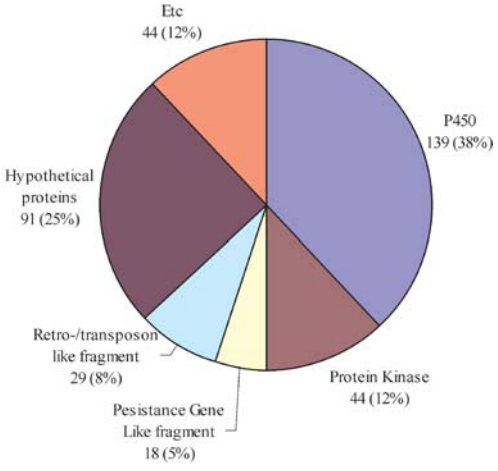

Fig. 4 Summary of a homology search across different plant species using the 365 fragments amplified from tomatoes with 15 PBA primer sets derived from mammalian P450 genes

that these primer-sets both amplified and detected polymorphisms that are associated with different functional regions in plant genomes.

Table 5 provides the representative summary on the association between the tomato PCR products amplified from the primers used in this research and the known *Arabidopsis* P450 genes. Some PCR products indicated relatively high homology (>70%) with the P450 genes and the sequences distributed among the five chromosomes of *Arabidopsis*.

Applicability of P450 gene-analogues

Figure 5 summarizes, in schematic form, a profile of the P450 primer annealing based on 365 fragments amplified by 15 primer sets from six different tomato genotypes. Fragments were generated in this study using all possible primer combinations: forward and reverse (F-R), forward and forward (F-F) and reverse and reverse (R-R). This could suggest that the primers amplified not only specific gene fragments but also multiple-analogues in the same PCR amplification, similar to RAPD.

Discussion

Pitfalls in plant genetic diversity studies

Analyses of plant genetic diversity at the molecular level have been conducted using many molecular markers such as isozymes, RFLPs, RAPDs, AFLPs, SSRs, ISSRs, etc. (Powell et al. 1996, Karp et al. 1997). These markers have both advantages and disadvantages but produce reliable

Table 5 A representative summary of the association survey between tomato PCR products amplified with PBA primer pairs and plant P450 genes in *Arabidopsis* that showed high homology (>70%) to the fragments. References from http://www.arabidopsis.org/info/genefamily/P450_functions.html

Identification of sequenced-tomato PCR clone	PBA primer pair that amplified the tomato clone in the left column	Plant CYPs Proteins deduced the from sequence	Putative activity of the plant CYP	Tomato sequence in the *Arabidopsis* chromosome
223-A1FhB6-1	CYP1A1F/heme 2B6	CYP51A2	Obtusifoliol 14alpha-demrthlase	I
223-A1FhC19-3	CYP1A1F/heme 2C19	CYP86A1	Omega-hydroxylase for saturated and unsaturated C12 to C18 fatty acids	V
223-A1FhC19-3	CYP1A1F/heme 2C19	CYP86A8	Omega-hydroxylase for saturated and unsaturated C12 to C18 fatty acids	II
Momo-B6FB6R-2	CYP2B6F/CYP2B6R	CYP90A1	23 alpha-hydroxylase for 6-oxo-cathasterone	V
Momo-B6FB6R-2	CYP2B6F/CYP2B6R	CYP90B1	22s alpha-hydroxylase for 6-oxo-cathasterone	III
Momo-B6FhB6-7	CYP2B6F/heme 2B6	CYP79A2	N-hydroxylase for phenylalanine	V
Momo-B6FhB6-7	CYP2B6F/heme 2B6	CYP79B2	N-hydroxylase for tryptophan, tryptophan analogues	IV
Momo-B6FhB6-7	CYP2B6F/heme 2B6	CYP79B3	N-hydroxylase for tryptophan	II
223-B6FhC19-1	CYP2B6F/heme 2C19	CYP83A1	N-hydroxylase for indole-3-acetyldoxime	IV
223-B6FhC19-1	CYP2B6F/heme 2C19	CYP83B1	N-hydroxylase for indole-3-acetyldoxime	IV
209-C19FC19R-2	CYP2C19F/CYP2C19R	CYP88A3	Multifunctional entkaurenoic acid oxidase	I
209-C19FC19R-2	CYP2C19F/CYP2C19R	CYP88A4	Multifunctional entkaurenoic acid oxidase	II
209-C19FC19R-2	CYP2C19F/CYP2C19R	CYP701A3	Multifunctional entkaurenoic acid oxidase	V
Momo-C19FhB6-5	CYP2C19F/heme 2B6	CYP75B1	3′-hydroxylase for narigenin	V
223-C19FhC19-3	CYP2C19F/heme 2C19	CYP72B1	26-hydroxolase for brassinolide	II

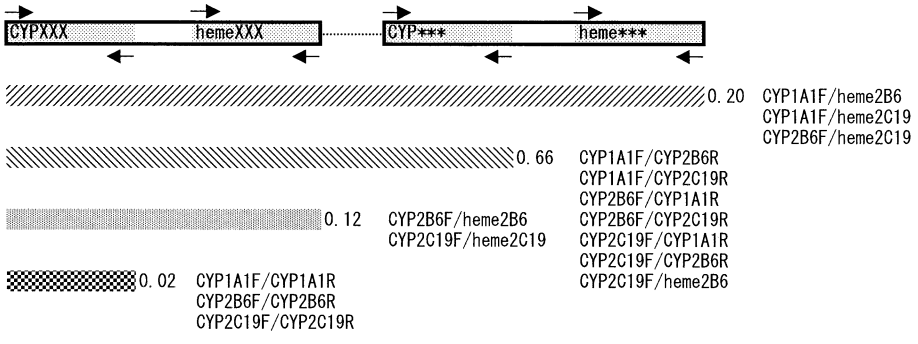

Fig. 5 Summary of the PBA primer annealing sites: sequence analyses of 365 PCR fragments from six tomato genotypes amplified with 15 PBA primer sets derived from mammalian P450 genes

information concerning genetic diversity, particularly in genetically neutral regions (Brown and Kresovich 1996; Karp 2002). The criteria that are usually applied in selecting specific markers include the speed and ease in processing information, and the cost-efficiency and reproducibility in addition to the quantity and type of genetic information that will be obtained (Karp et al. 1997). There was a continuing concern, however, as to whether measurement of genetic diversity by using such genetic markers truly reflects genome-wide variation. It was considered highly possible, therefore, that information obtained by using neutral markers, such as RAPDs, pertains to regions of the genome with an inherent bias (Monna et al. 1994), and thus does not reflect the entire genome. Also, it is known that SSRs tend to cluster at centromeric regions in plant varieties such as tomato (Solanaceae Genomics Network web site, http://www.sgn.cornell.edu/home_page.html) and barley and wheat (GrainGenes web site, http://wheat.pw.usda.gov/ggpages/maps.shtml).

Single Nucleotide Polymorphisms (SNPs) in commonly known genes for basic plant functions may therefore be of interest to address the above concerns and for general use as functional genomic markers (Nasu et al. 2002; Wheat SNPs web site, http://wheat.pw.usda.gov/ITMI/2002/WheatSNP.html). However, available genetic information, even on model species such as rice, is still meagre, and the application of SNP technology to diverse plant species may consequently be unworkable. Additionally, little is yet known concerning the diversity of functional regions in the genome.

Data mining over the *Arabidopsis* genome on the specific multi-gene family

To further address these questions, we utilised the gene-analogue genetic diversity of the human P450 gene. An enormous amount of genetic information on this gene family has so far been compiled, which is in stark contrast to the dearth of comparable plant information. From the

results of our data mining of the *Arabidopsis* genome using human P450 sequences, it was revealed that the primer sequences we used were distributed throughout the *A. thaliana* genome (Fig. 1), and small fragments corresponding to the original cDNA sizes were also obtained (Figs. 2a, b). Thus, the results of our initial trials to test whether the sequence information of mammalian P450 genes can be utilised to analyse functional genetic diversity in higher plant species were very positive.

Rapid and efficient assessment of polymorphisms, and genetic diversity via PCR

From the results of our PCR analysis, P450 gene-like fragments showed polymorphisms both between and within plant species, suggesting that these markers were indeed effective tools for estimating genetic diversity in a range of plant species.

Annotation of the *A. thaliana* genome has shown that about 0.9% of 29,000 genes (272 genes and 26 pseudogenes) could be categorised as cytochrome P450 genes: (http://arabidopsis-p450.biotec.uiuc.edu/; http://arabidopsis.org/info/genefamily/p450.html; http://drnelson.utmem.edu/CytochromeP450.html; http://www.biobase.dk/p450/; Riechmann et al. 2001). The results of our data mining, and the subsequent sequence analyses of the PCR products obtained from the plant species, showed that the primers used in this study could amplify not only specific genes but also multiple analogues of the cytochrome P450 gene family (Fig. 5). It should be emphasized that only 15 pairs consisting of eight primers provided sufficient polymorphism data in a highly reproducible manner (Tables 3 and 4). This has enormous advantages over conventional RAPD markers that require dozens of primers to identify an adequate number of polymorphisms in a reproducible fashion (Williams et al. 1990). Furthermore, the sequences of the amplified fragments were homologous to a variety of proteins (Fig. 3).

Based on Table 5, it was suggested that annotated/putative P450 or analogues would be obtained with the PBA primers, at least partial sequences of them would be proof of the hypothesis and the possibility for the application to genome-wide marker sets. However, not all sequences were associated with the plant P450 (Fig. 3), but at least some of the amplified products could be regarded as the plant P450-associated fragment; and also other fragments that have relationships with the reported or annotated genes would be of use for the functional genomic markers for the genetic-diversity study. A provisional proof mapping of the clones listed in Table 5 corresponded to the linkage position of the known P450 genes (Yamanaka et al., in preparation).

Overall, our findings strongly indicate that PBA markers are a valuable new tool for analysing both diversity and conservation in plant genomes, with the need for only a relatively small numbers of markers. Proof mapping of the amplified clones that were sequenced, is currently being undertaken in our laboratory for tomato, potato and *Poaceae*, and provisional mapping-results indicates dispersal of P450 analogue DNA clones over these genomes (Yamanaka et al., in preparation).

Functional genomic information for genetic diversity study

Obtaining primary information about plant genetic diversity is important fundamental work in genetic conservation. Both in situ and ex situ (genebank) management, and marker sets for genome-wide functional analysis, have become effective tools in analysing and utilising unexploited gene pools (Tanksley and McCouch 1997; Watanabe and Iwanaga 1999). It was previously suggested that information from multi-gene families, such as cytochrome P450, could be used as a benchmark in functional genomic analysis and also for diverse applications in genetic studies (Somerville and Somerville 1999). We have demonstrated that P450 gene-analogues are as effective as new genetic markers for genetic diversity studies in plants, which reflects both functional and genome-wide regions. We designate these markers as PBAs (P450 based analogues).

To employ genetic markers with a genome-wide distribution, a number of transposons and retrotransposons have previously been characterised and successfully used for the estimation of genetic diversity and the determination of phylogenetic relationships (e.g. Deininger 1989). Because our new assays are PCR-based however, the cytochrome P450 gene sequences used in this study would be more useful for a rapid, simple and precise evaluation of genetic diversity, as transposon and retrotransposon markers require Southern hybridisation and/or sequencing procedures. It should also be noted that our new assays are applicable not only for model plants, that have been well-studied, but also for the species without ample sequence or genetic map information. This will prove to be extremely useful for the evaluation of primary genetic information in less well-studied, but potentially interesting and valuable, plant species (Watanabe et al., in preparation). These new markers have now been employed by us for the evaluation of diverse species such as potato (Yamanaka et al., in preparation), sweetpotato (*Ipomoea*), Japanese arrow-head (*Sagittaria* spp.) (Tanaka et al. 2001; Watanabe et al., in preparation), taro (*Colocasia*), rice (Yamanaka et al. 2003; Yamanaka et al., in preparation) and various endemic wild-plant species. We hope to provide novel data on these and other species in the near future.

Acknowledgements This work was supported by grants from MEXT, Japan: RFTF-96-L603 and RFTF 00L01602 (K.N.W. and S.Y.). K.N.W. thanks the Sumitomo Foundation also for its financial support. The authors thank Prof. T. Fujimura, University of Tsukuba, for his critical review of our draft manuscript.

References

Bousquet J, Simon L, Lalonde M (1990) DNA amplification from vegetative and sexual tissues of the tree using the polymerase chain reaction. Can J For Res 20:254–257

Brown SM, Kresovich S (1996) Molecular characterization for plant genetic resources conservation. In: Paterson AH (ed) Genome mapping in plants. Landes, Texas, pp 85–93

Deininger PL (1989) SINEs: short interspersed repeated DNA elements in higher eucaryotes. In: Berg DE, Howe MM (eds) Mobile DNA. Am Soc Microbiol, Washington D.C., pp 619–636

Donaldson RP, Luster DG (1991) Multiple forms of plant cytochromes P-450. Plant Physiol 96:669–674

Inui H, Komada T, Ohkawa Y, Ohkawa H (2000) Herbicide metabolism and cross-tolerance in transgenic potato plants co-expressing human *CYP1A1, CYP2B6* and *CYP2C19*. Pestic Biochem Physiol 66:116–129

Jarvis D, Hodgkin T (1999) Wild relatives and crop cultivars detecting natural introgression and farmer selection of new genetic combination in agroecosystems. Mol Ecol 8:S159–S173

Karp A (2002) The new genetic era: will it help us in managing genetic diversity? In: Engels JMM, Rao VR, Brown AHD, Jackson MT (eds) Managing plant genetic diversity. Wallingford, CAB Publishing, pp 43–56

Karp A, Kresovich S, Bhat KV, Ayad WG, Hodgkin T (1997) Molecular tools in plant genetic resources conservation: a guide to the technologies. International Plant Genetic Resources Institute, Rome

Kessmann H, Choudhary, AD, Dixon, A (1990) Stress responses in alfalfa *(Medicago sativa* L.). III. Induction of medicarpin and cytochrome P450 enzyme activities in elicitor-treated cell suspension cultures and protoplasts. Plant Cell Rep 9:38–41

Kiyokawa S, Ohbayashi M, Shimada Y, Kikuchi Y (1997) PCR-amplification of sequences encoding the heme-binding region of plant cytochrome P450. Plant Biotech 14:175–178

Monna L, Miyao A, Inoue T, Fukuoka S, Yamazaki M, Zhong HS, Sasaki T, Minobe Y (1994) Determination of RAPD markers in rice and their conversion into Sequence Tagged Sites (STSs) and STS-specific primers. DNA Res 1:139–148

Nasu S, Suzuki J, Ohta R, Hasegawa K, Yui R, Kitazawa N, Monna L, Minobe Y (2002) Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (*Oryza sativa*, *Oryza rufipogon*) and establishment of SNP markers. DNA Res 9:163–171

Ohkawa H, Imaishi H, Shiota N, Yamada T, Inui H, Ohkawa Y (1998) Molecular mechanisms of herbicide resistance with special emphasis on cytochrome P450 mono-oxigenases. Plant Biotech 15:173–176

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol Breed 2:225–238

Riechmann JL, Heard J, Martin G, Reuber L, Jiang C-Z, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ, Ghandehari D, Sherman BK, Yu G-L (2001) *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. Science 290:2105–2110

Shalk M, Nedelkina S, Schoch G, Batard Y, Werck-Reichhart D (1999) Role of unusual amino-acid residues in the proximal and distal heme regions of a plant P450, CYP73A1. Biochemistry 38:6093–6103

Somerville C, Somerville S (1999) Plant functional genomics. Science 285:380–383

Song WC, Funk CD, Brash AR (1993) Molecular cloning of an allene oxide synthase: a cytochrome P450 specialized for the metabolism of fatty acid hydroperoxides. Proc Natl Acad Sci USA 90:8519–8523

Suzuki E, Watanabe JA, Watanabe KN (2001) Possibility of genetic diversity study in plants using sequence information of mammalian xenobiotics metabolizing cytchrome P450 genes (In Japanese). Breed Res 3 Suppl (1):187

Tanaka M, Watanabe JA, Watanabe KN (2001) Presence of genetic variation in the endangered semi-cultivated species, Suita Kuwai (*Sagittaria triforia* L. *var suitensis*). Breed Res 3 (Suppl 1): 162

Tanksley SD, McCouch SR (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. Science 277:1063–1066

Teutsch HG, Hasenfratz MP, Lesot A, Stoltz C, Garnier JM, Jeltsch JM, Durst F, Werck-Reichhart D (1993) Isolation and sequence of a cDNA encoding the Jerusalem artichoke cinnamate 4-hydroxylase, a major plant cytochrome P450 involved in the general phenylpropanoid pathway. Proc Natl Acad Sci USA 90:4102–4106

Wagner DB, Furnier GR, Saghai-Maroof MA (1987) Chloroplast DNA polymorphism in lodgepole and jack pines and their hybrids. Proc Natl Acad Sci USA 84:2097–2100

Watanabe KN, Iwanaga M (1999) Plant genetic resources and its global contribution. Plant Biotech 16:7–13

Watanabe KN, Takeda Y, Inui H, Kasai K, Ohkawa H (2000) Survey of xenobiotic-metabolizing cytochrome P450 homologues in higher plant species. Memo Inst BOST, Kinki Univ 5:9–18

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18:6531–6535

Yamanaka S, Nakamura I, Sato Y-I, Watanabe KN (2003) Evaluation on genetic resources based on the diversity of functional regions. 1. Genetic diversity of wild rice estimated by variations on xenobiotics-metabolizing cytochrome P450 gene-analogues. Breeding Res 5 (Suppl 1):249